



**University of  
Zurich**<sup>UZH</sup>

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2014

---

## **Integration of spoken and written words in beginning readers: A topographic ERP study**

Jost, Lea B ; Eberhard-Moscicka, Aleksandra K ; Frisch, Christine ; Dellwo, Volker ; Maurer, Urs

**Abstract:** Integrating visual and auditory language information is critical for reading. Suppression and congruency effects in audiovisual paradigms with letters and speech sounds have provided information about low-level mechanisms of grapheme-phoneme integration during reading. However, the central question about how such processes relate to reading entire words remains unexplored. Using ERPs, we investigated whether audiovisual integration occurs for words already in beginning readers, and if so, whether this integration is reflected by differences in map strength or topography (aim 1); and moreover, whether such integration is associated with reading fluency (aim 2). A 128-channel EEG was recorded while 69 monolingual (Swiss)-German speaking first-graders performed a detection task with rare targets. Stimuli were presented in blocks either auditorily (A), visually (V) or audiovisually (matching: AVM; nonmatching: AVN). Corresponding ERPs were computed, and unimodal ERPs summated ( $A + V = \text{sumAV}$ ). We applied TANOVAs to identify time windows with significant integration effects: suppression ( $\text{sumAV} - \text{AVM}$ ) and congruency ( $\text{AVN} - \text{AVM}$ ). They were further characterized using GFP and 3D-centroid analyses, and significant effects were correlated with reading fluency. The results suggest that audiovisual suppression effects occur for familiar German and unfamiliar English words, whereas audiovisual congruency effects can be found only for familiar German words, probably due to lexical-semantic processes involved. Moreover, congruency effects were characterized by topographic differences, indicating that different sources are active during processing of congruent compared to incongruent audiovisual words. Furthermore, no clear associations between audiovisual integration and reading fluency were found. The degree to which such associations develop in beginning readers remains open to further investigation.

DOI: <https://doi.org/10.1007/s10548-013-0336-4>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-87938>

Journal Article

Published Version

Originally published at:

Jost, Lea B; Eberhard-Moscicka, Aleksandra K; Frisch, Christine; Dellwo, Volker; Maurer, Urs (2014). Integration of spoken and written words in beginning readers: A topographic ERP study. *Brain Topography*, 27(6):786-800.

DOI: <https://doi.org/10.1007/s10548-013-0336-4>

# Integration of Spoken and Written Words in Beginning Readers: A Topographic ERP Study

Lea B. Jost · Aleksandra K. Eberhard-Moscicka ·  
Christine Frisch · Volker Dellwo · Urs Maurer

Received: 4 July 2013 / Accepted: 15 November 2013 / Published online: 24 November 2013  
© Springer Science+Business Media New York 2013

**Abstract** Integrating visual and auditory language information is critical for reading. Suppression and congruency effects in audiovisual paradigms with letters and speech sounds have provided information about low-level mechanisms of grapheme-phoneme integration during reading. However, the central question about how such processes relate to reading entire words remains unexplored. Using ERPs, we investigated whether audiovisual integration occurs for words already in beginning readers, and if so, whether this integration is reflected by differences in map strength or topography (aim 1); and moreover, whether such integration is associated with reading fluency (aim 2). A 128-channel EEG was recorded while 69 monolingual (Swiss)-German speaking first-graders performed a detection task with rare targets. Stimuli were presented in blocks either auditorily (A), visually (V) or audiovisually (matching: AVM; nonmatching: AVN). Corresponding ERPs were computed, and unimodal ERPs summated ( $A + V = \text{sumAV}$ ). We applied TANOVAs to identify time windows with significant integration effects: suppression ( $\text{sumAV} - \text{AVM}$ ) and congruency ( $\text{AVN} - \text{AVM}$ ). They were further characterized using GFP and 3D-centroid analyses, and significant effects were

correlated with reading fluency. The results suggest that audiovisual suppression effects occur for familiar German and unfamiliar English words, whereas audiovisual congruency effects can be found only for familiar German words, probably due to lexical-semantic processes involved. Moreover, congruency effects were characterized by topographic differences, indicating that different sources are active during processing of congruent compared to incongruent audiovisual words. Furthermore, no clear associations between audiovisual integration and reading fluency were found. The degree to which such associations develop in beginning readers remains open to further investigation.

**Keywords** Audiovisual integration · Language processing · ERP mapping · Children · Reading

## Introduction

Linking the visual form of words to the known phonological and semantic properties of spoken words is a key element of reading. A crucial step at the beginning of learning to read is thus establishing links between print and spoken language, and by that, developing new audiovisual grapheme-phoneme associations (Blomert 2011). Linking print to speech is not only important for normal processes of learning to read. It may also be involved in the development of reading disorders such as dyslexia due to reduced orthographic-phonological connectivity (Blomert 2011; Wimmer and Schurz 2010).

Given the central function of linking print to speech, especially in the first years of reading acquisition, when children are at the initial stages of learning correspondences between graphemes and phonemes, research on this

L. B. Jost · A. K. Eberhard-Moscicka · C. Frisch ·  
U. Maurer (✉)  
Division of Cognitive Neuroscience, Department of Psychology,  
University of Zurich, Binzmuehlestr. 14/4, 8050 Zurich,  
Switzerland  
e-mail: urs.maurer@psychologie.uzh.ch

L. B. Jost · A. K. Eberhard-Moscicka · U. Maurer  
Neuroscience Center Zurich, University of Zurich and ETH  
Zurich, Zurich, Switzerland

V. Dellwo  
Phonetics Laboratory, University of Zurich, Zurich, Switzerland

topic has gained substantial interest. One way to address the question of how print is linked to speech is to investigate the integration of visual and auditory language information by using audiovisual paradigms. The rationale behind this approach is that visual language code is transformed into phonological and semantic language code during reading. Interactions between visual and auditory language input in audiovisual experiments thus indicate print to speech conversion in reading.

Audiovisual paradigms have been widely used in research about mechanisms of multisensory integration (for a review see e.g., Murray and Spierer 2009) and in research about audiovisual speech perception (e.g., Klucharev et al. 2003), but have only rarely been applied to investigate reading, as reviewed below.

A typical audiovisual experiment is composed of four conditions: stimuli are presented either auditorily (A), visually (V) or audiovisually (matching: AVM; non-matching: AVN). Within this experimental framework two main approaches can be applied in order to investigate audiovisual integration mechanisms. The first approach is pursued by contrasting the sum of responses to auditory and visual stimuli (sumAV) to the responses to audiovisual stimuli (AV) (Raij et al. 2000). If auditory and visually presented stimuli were processed independently, then the neural responses induced by an audiovisual stimulus should be equal to the sum of the responses to the unimodal presented stimuli (additive model, Besle et al. 2004; Giard and Peronnet 1999). Any neural activity that differs from the mere summation may reflect interactions between auditory and visual processing (Besle et al. 2004), especially if these differences occur early during stimulus processing in trials without motor response (Besle et al. 2004, 2009; Cappe et al. 2010). Such interaction has been shown to result in a reduced response strength at temporal sensors for audiovisual compared to summated unimodal stimuli in an MEG study ( $AV < \text{sumAV}$ ) (Raij et al. 2000), and was interpreted as suppressive interaction. Following these authors, we will be referring to the results of the AV versus sumAV contrast as audiovisual ‘suppression effect’.

The second approach to investigate audiovisual integration mechanisms is pursued by comparing neural activation in response to bimodal matching stimuli with neural activation in response to bimodal non-matching stimuli (Blau et al. 2010; van Atteveldt et al. 2004). According to van Atteveldt et al. (2007a, b) this congruency effect indicates multisensory integration, since a discrepancy between matching and nonmatching letters and speech sounds cannot be established without the unisensory inputs having been integrated successfully.

Both these approaches have been used in several neuroimaging studies using either spatially or temporally

sensitive methods to investigate integration of letters and speech sounds.

An MEG study in adults showed a suppression effect in neural activation around 350 ms, when the audiovisual matching condition was compared to the sum of the two unimodal conditions. This suppression effect was localized in the superior temporal sulcus (STS), and interpreted as reflecting optimization processes following learning (Raij et al. 2000). Another MEG study reported increased oscillations for congruent relative to incongruent audiovisual stimuli during the first 250 ms in left auditory cortex regions, and decreased oscillations between 250 and 500 ms in bilateral visual cortex regions (Herdman et al. 2006) suggesting early crossmodal interaction in auditory language regions.

Studies investigating letter-speech sound associations using spatially sensitive methods showed that the unimodal auditory association cortex (planum temporale, PT) plays a role in the integration process (van Atteveldt et al. 2004, 2007a). This response from the auditory cortex was reflected by an enhanced response to congruent as opposed to a suppressed response to incongruent letter-speech sound pairs (van Atteveldt et al. 2004, 2007a). Contrasting audiovisual matching to the sum of the two unimodal conditions revealed activation that was superadditive in the PT ( $AV > \text{sumAV}$ ), but subadditive in the STS/STG ( $AV < \text{sumAV}$ ) (van Atteveldt et al. 2004). According to these authors speech processing in PT is modulated by visual input, whereas the STS/STG regions are integration sites where auditory and visual information converge (van Atteveldt et al. 2004).

While fMRI studies using single letters and phonemes reported that audiovisual interaction is reflected by larger activation for congruent than incongruent stimuli (van Atteveldt et al. 2004, 2007a), a study using entire words found the reverse effect (Hocking and Price 2009). Activation was larger for incongruent than congruent stimuli across the whole network of regions, including the STS/STG, that were involved in processing audiovisual stimuli. This suggests that entire visual and auditory words are integrated differently compared to single letters and phonemes.

Another experimental approach to investigate how print is linked to speech is to use an auditory oddball paradigm in combination with visual stimuli. The auditory mismatch negativity (MMN) has been shown to increase when paired with visual letters compared to auditory presentations alone (Froyen et al. 2008) or compared to scrambled letter pairings (Mittag et al. 2013). This suggests that visual input modulates auditory processing within 200 ms of stimulus presentation.

While all these studies shed light on audiovisual integration of speech and letters in adult skilled readers, little is

known how print and spoken language is linked when children learn to read. A developmental study using an audiovisual paradigm with letters and speech sounds revealed that 9 year-old children already show an adult-like congruency effect in the planum temporale (Blau et al. 2010). However, the extent of this effect seemed to be reduced compared to adults, suggesting protracted development of letter-sound associations over the course of reading acquisition. In accordance, it appears to take children several years to show a similar visual enhancement of the MMN as adults (Froyen et al. 2009). Whereas 11 year-olds showed an immature enhancement effect compared to the adults, the 8 year-olds did not show any effect after 1 year of reading instruction.

The development of these audiovisual integration effects seems not only to depend on the age of the children, but also on the level of their reading skills, as indicated by dyslexia studies. It has been shown that in contrast to fluently reading children, dyslexic children show little or no activation increase for congruent compared to incongruent letter-speech sound pairs in the planum temporale/Heschl sulcus and the STS (Blau et al. 2010). Similarly, dyslexic children did not show any early visual enhancement of the MMN even after 4 years of reading instruction (Froyen et al. 2011). These studies pointing towards an important role of integrating orthographic and phonological representations are in agreement with recent reading theories (Blomert 2011; Wimmer and Schurz 2010).

Even though the discussed studies with letters and speech sounds have been informative about the mechanisms of audio-visual integration during reading, only little is known about the role of such multisensory processes related to entire word reading. In contrast to single graphemes, words may be read by converting groups of letters or even entire word forms to pronunciations while at the same time activating meaning (for a review see Ehri 2005). Moreover, as reviewed above, only scarce knowledge exists regarding the development of audiovisual integration and the role it plays for early reading acquisition.

In this paper, we present data from 69 monolingual (Swiss-) German speaking children in first grade. The participants performed a detection task with familiar German word and unfamiliar English word stimuli which were presented either auditorily (A), visually (V) or audiovisually (matching: AVM; nonmatching: AVN). In order to enable new insights into the development of audiovisual integration and its role in reading fluency, audiovisual integration mechanisms were studied applying the two paradigms used in previous research: “suppression effect” (sumAV vs. AVM) and “congruency effect” (AVN vs. AVM). This study pursued two aims.

**Aim 1:** In a first step, we investigated if audiovisual integration occurs for words already at the initial stages of

reading acquisition which would indicate that processes of linking print to speech occur, even though reading is presumably not automatized yet. Moreover, including familiar German and unfamiliar English words allowed to test whether audiovisual integration depends on lexical-semantic properties of the stimuli. To this end, we searched for time windows of audiovisual integration for familiar German and unfamiliar English words, applying a time-point-to-time-point topographic analysis of variance (TANOVA). This allowed us to detect differences between two conditions, i.e. between AVM versus AVN (congruency effect) and between sumAV and AVM (suppression effect). If such integration could be observed, we further investigated how it manifests in ERPs, that is whether audiovisual integration is reflected by differences in topography or map strength. The distinction between GFP and topographic effects is essential, as different topographies directly indicate different source configurations, whereas different GFP in the absence of a topographic effect indicates different activation strength of the same sources (Michel et al. 2004).

**Aim 2:** Based on studies suggesting reduced audiovisual integration of letters and speech sounds in dyslexic children (Blau et al. 2010; Froyen et al. 2011) we further wanted to test whether audiovisual integration is associated with reading fluency. Considering that building new connections between written and spoken language is a crucial step in reading acquisition (Blomert 2011), we expected more fluent readers to show stronger audiovisual integration than slow readers. Investigating a potential association was possible due to the large group of children with a wide range of reading skills.

## Materials and Methods

### Participants

We report data from 69 monolingual (Swiss-)German children (mean age 7.6 years, range 6.7–8.5 years, 27 girls, 5 left-handed) who were tested at the end of first grade. From an original group of 70 subjects, one subject was excluded due to low signal-to-noise ratio in the ERPs. All subjects had normal or corrected-to-normal vision and a nonverbal IQ of 80 or above (block design, Petermann and Petermann 2010). They were screened for a history of neurological diseases or psychiatric disorders and the parents rated their children’s behavior by the child behavior checklist (CBLC, Achenbach 1991). All subjects were part of a larger longitudinal study about learning English as a foreign language. The data reported here were collected before the children had started learning English. The parents gave informed consent and the children were

compensated with a book voucher of 40 CHF for their participation. In addition, the parents were given a brief written report about the behavioral session of their child. The study protocol was in agreement with the local ethics committee.

### Procedure

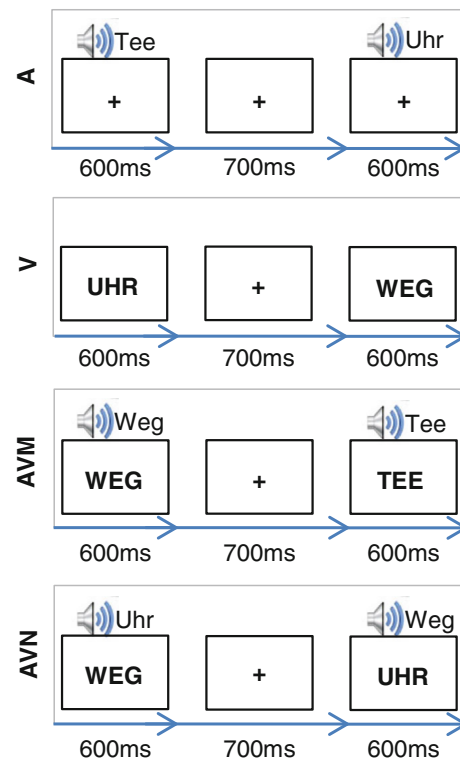
All children participated in two sessions consisting of one behavioral session and one EEG-session. The behavioral assessment took about 1.5 h per child and was accomplished either in a room provided by the child's school, at the child's home or at the Department of Psychology at the University of Zurich. The mobile EEG system allowed us to do recordings either in extra rooms provided by some schools, or in our laboratory at the Department of Psychology at the University of Zurich. Before doing the recordings at schools, a standard quality check was applied in order to ensure for the absence of 50 Hz noise. The EEG-session took about 3.5 h per child, including breaks between experiments.

### Behavioral Assessment

Behavioral measures assessed different aspects of language processing such as word and pseudoword reading fluency (SLRT-I, Landerl et al. 2006; Moll and Landerl 2010), vocabulary knowledge (vocabulary subtest, HAWIK-IV, Petermann and Petermann 2010), rapid automatized naming (RAN), reading comprehension (SLS 1-4, Mayringer and Wimmer 2003), phonological awareness (BAKO, Stock et al. 2003) and auditory memory span (digit span subtest, HAWIK-IV, Petermann and Petermann 2010). Non-verbal IQ (block design subtest, HAWIK-IV, Petermann and Petermann 2010) was used as an exclusion criterion ( $IQ < 80$ ).

### Task

The audiovisual experiment was one of several short experiments, which were pseudo-randomized in order of presentation. An EEG was recorded during this experiment while children performed a detection task with a rare target among word stimuli (see below). The words presented were 10 familiar German words (e.g. *WEG*), 10 unfamiliar simple English words (pronunciation according to German grapheme-phoneme correspondence rules and phonetic inventory, e.g. *PEN*) and 10 unfamiliar complex English words (pronunciation not according to German grapheme-phoneme correspondence rules or phonetic inventory, e.g. *EYE*). The number of items was limited to 10 per word list due to the limited sample of English words the children were expected to know at the follow-up session after 1 year of learning



**Fig. 1** Experimental design: German word and unfamiliar English word stimuli were presented either auditorily (A), visually (V), audiovisually matching (AVM) or audiovisually nonmatching (AVN)

English (not part of this study, which focuses on the children before learning English), and due to the matching procedure between the three word lists (see below). This analysis focuses on familiar German words and unfamiliar simple English words. After the EEG-session, each child's knowledge of the simple English words was tested. Accordingly, they knew only 1.68 words in average (16.8 %) in written or spoken form. The English words a child knew were excluded on an individual basis from the computation of the ERPs, to make sure that only English words were included which did not have lexical-semantic content.

The lists of familiar German words and unfamiliar simple English words were matched for number of letters, word type, word length, number of spoken syllables and frequency of letters. Moreover, the two lists did not differ regarding bigram frequency, nor regarding number of neighbors based on a German text corpus (CELEX Lexical Database, Baayen et al. 1993; implemented in WordGen application, Duyck et al. 2004). The visual stimuli were presented for 600 ms in black (Arial, bold, font size 28, uppercase letters) on a light-grey background in the center of the screen. The auditory stimuli, spoken by a German-English bilingual male speaker, were all scaled to the same length (600 ms) using Praat software (Boersma 2001).

The experiment was divided into two parts, each of which took about 12 min. Children were allowed to take a



short break after 6 min. The stimuli were presented either only auditorily (A), visually (V) or audio-visually (AV). In the bimodal condition (AV) the stimuli were either matching (AVM) or nonmatching (AVN) (see also Fig. 1). There were 80 trials for each of the 12 stimulus types (4 modalities  $\times$  3 word types). Each word was presented 24 times in the visual (8 unimodal V, 8 bimodal matching, 8 bimodal nonmatching) and 24 times in the auditory modality (8 unimodal A, 8 bimodal matching, 8 bimodal nonmatching). Given the overlap in the audiovisual matching condition, there were 40 trials when the same word appeared either in the visual or auditory modality.

All in all, 1,056 stimuli were presented, of which 96 were target stimuli. The duration for both visually and auditorily presented stimuli was 600 ms followed by a 700 ms interstimulus interval (ISI). The different stimulus conditions (A, V, AVM and AVN) and the word lists (German, simple English, complex English) were presented in blocks. Block order was counterbalanced across subjects. The participants were asked to press a device button with their preferred hand after the rare target stimulus *PINK*, which is a common word in German and in English. The target stimuli were presented visually, auditorily, or audiovisually according to the modality blocks at a ratio of 9 %. In the bimodal conditions, targets also appeared only visually or auditorily to prevent purely visual or purely auditory strategies. Moreover, as there was only one target, audiovisually presented targets were always matching.

To verify that only children who attended and responded to the stimuli were included in the analyses, a minimum overall accuracy of 0.50 correct responses served as inclusion criteria. None of the children needed to be excluded because of low attention during the EEG.

#### Electrophysiological Recording and Analysis

The 128-channel EEG (HydroCel, Electrical Geodesics Inc.) was recorded against Cz-reference at 250 Hz/channel with a high pass filter of 0.1 Hz and a low pass filter of 100 Hz. Impedance was kept below 50 k $\Omega$ . ERPs were processed in Vision Analyzer software (Brain Products GmbH) including digital filtering (lowpass: 30 Hz, 48 dB/oct; highpass: 0.3 Hz, 24 dB/oct), interpolating bad channels and correcting for vertical and horizontal eye movement artifacts using an independent component analysis (Jung et al. 2000). The continuous EEG was transformed to the average reference (Lehmann and Skrandies 1980). Before averaging, the remaining trials were epoched 150 ms prior and 1,320 ms following the stimulus onset. Trials with artifacts exceeding 180  $\mu$ V difference between maximum and minimum in any channel were automatically rejected. All ERPs were corrected for a constant 20 ms delay, as revealed by a timing test, and baseline corrected

(−150 to 0 ms). The corrected epoch length used for the analysis was from −170 to 1,300 ms. Averaging was done separately for each word category (familiar German words, unfamiliar English words) and modality (A, V, AVM, AVN) including only nontarget stimuli. Unimodal ERPs were summated (sumAV). Difference ERPs between conditions of interest (sumAV vs. AVM; AVN vs. AVM) were computed individually, before computing grand averages of ERPs and ERP differences.

#### Statistical Analysis

For the analysis of the resulting ERPs we used a two-step strategy (Maurer et al. 2008, 2010). In order to investigate if audiovisual integration occurs for familiar German words and unfamiliar English words already at the initial stages of reading acquisition (aim 1), we first tested for time windows reflecting audiovisual integration effects (suppression effect: sumAV vs. AVM and congruency effect: AVN vs. AVM). For this purpose we applied four topographic analyses of variance (TANOVA, part of the LORETA-KEY software package) across all subjects (separately for familiar German words/unfamiliar English words). These point-to-point TANOVAs were conducted on non-normalized (raw) maps to discover any systematic differences between the conditions of interest (sumAV vs. AVM; AVN vs. AVM) by running a nonparametric randomization test (Holmes et al. 1996) on the GFP of difference maps between two conditions (Lehmann and Skrandies 1980; Strik et al. 1998). To control for multiple comparisons across these analyses, a more conservative alpha level of  $p < 0.01$  in at least three consecutive time frames was adopted (Maurer et al. 2008, 2010). This TANOVA approach has been applied in several previous studies to identify time segments during which two conditions were processed differently (Maurer et al. 2003, 2008, 2010; Murray et al. 2004; Schulz et al. 2008), or during which changes over time occurred due to training (e.g., Stein et al. 2006). Given the scarce prior knowledge about the time of occurrence of effects, a TANOVA allowed us to detect whether and when audiovisual integration occurs without the need to preselect a subset of electrodes. Because TANOVA differences between non-normalized (raw) maps can result from differences in map strength or from differences in topography, we investigated the resulting time windows in more detail in a second step. Based on the average across the resulting significant time windows global field power (GFP) (Lehmann and Skrandies 1980) and 3D-centroids were computed and analyzed applying ANOVAs. GFP equals the root mean square across all recording electrodes and is a reference-independent measure of response strength (Lehmann and Skrandies 1980). The positive and negative 3D centroids are voltage-weighted means of the positions of all electrodes showing positive or negative values, respectively.

They are independent of GFP and thus a purely topographic measure based on all 129 electrodes. Centroid positions are defined in Talairach space by x- (left–right), y- (posterior–anterior), and z- (inferior–superior) coordinates (Maurer et al. 2010). ANOVAs for repeated measures were computed on GFP with the within subject factor “suppression” (sumAV vs. AVM), or “congruency” (AVN vs. AVM), respectively. Analogous ANOVAs on centroid positions additionally included the within subject factor “polarity” (positive vs. negative, reported in interaction effects) while the x-, y-, and z-axes were treated as multivariate dependent measures (Maurer et al. 2003). In an additional analysis we aimed at testing the association between reading fluency and audiovisual integration (aim 2). For this purpose, we computed correlations between reading fluency and the differences between the two conditions (sumAV–AVM, AVN–AVM, respectively) that revealed significant effects of audiovisual integration in the GFP and centroid ANOVAs from aim 1. Reading fluency, defined as correct words per minute, was based on the average of 3 z-transformed subtests of the SLRT-I (Landerl et al. 2006) and SLRT-II (Moll and Landerl 2010).

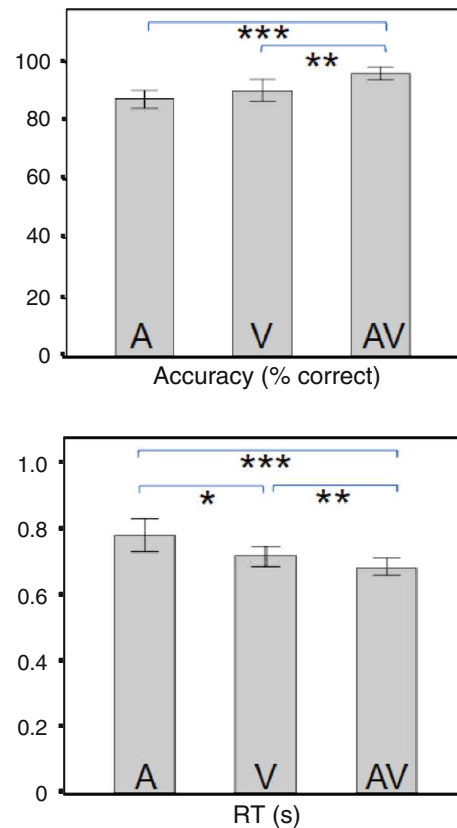
In addition to the ERP analyses of the non-target stimuli, we computed behavioral analyses for the responses to the targets. Behavioral accuracy (percentage of detected targets) and reaction time (RT to detected targets) were analyzed by computing two ANOVAs for repeated measures with within-subject factor “modality” (auditory, A; visual, V; audiovisual, AV). Statistical trends ( $p < 0.1$ ) are indicated in the results, but not further interpreted.

## Results

### Behavioral Data

Both ANOVAs on the behavioral accuracy and RT of the responses to the targets (averaged across familiar German words and unfamiliar English words) revealed significant main effects of modality (accuracy  $F(2,67) = 16.446$ ,  $p < 0.001$ ; RT  $F(2,67) = 11.064$ ,  $p < 0.001$ ).

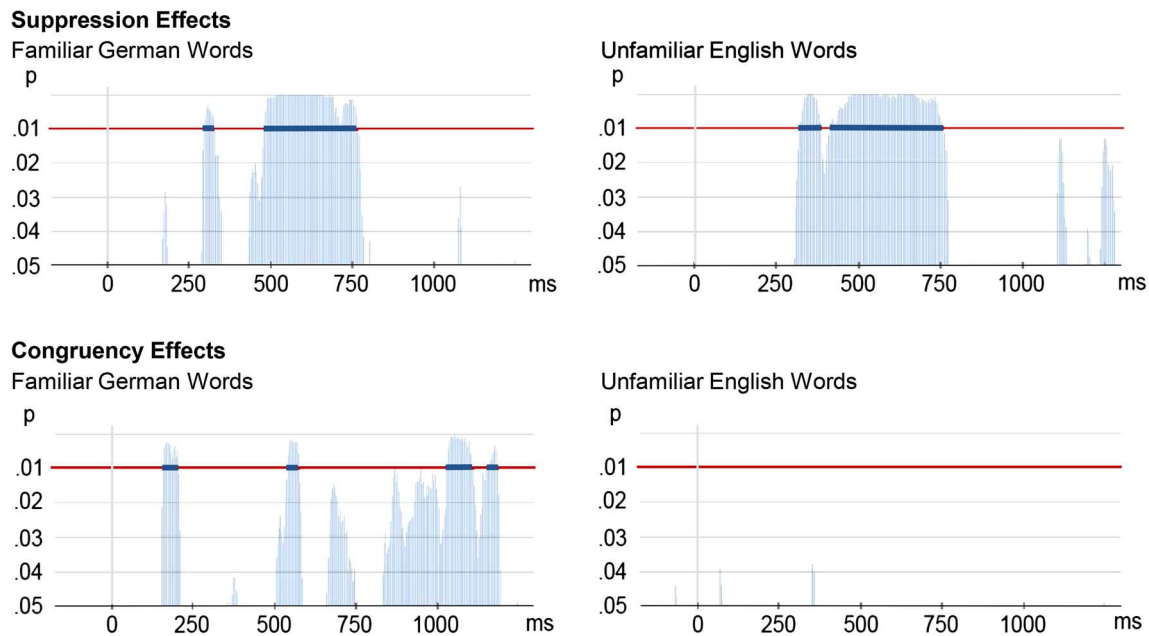
Post-hoc  $t$  tests showed that children detected targets more accurately and faster when they were presented simultaneously in both modalities compared to when they were presented only visually (accuracy:  $t(68) = 3.432$ ,  $p = 0.001$ ; RT:  $t(68) = 3.146$ ,  $p < 0.01$ ) or only auditorily (accuracy:  $t(68) = 5.310$ ,  $p < 0.001$ ; RT:  $t(68) = 3.900$ ,  $p < 0.001$ ). There was no difference in accuracy between auditorily and visually presented stimuli ( $t(68) = 1.481$ ,  $p = ns$ ). However, the RT was lower when the stimuli were presented visually compared to auditorily ( $t(68) = 2.488$ ,  $p < 0.05$ ) (see Fig. 2).



**Fig. 2** Mean accuracy (percentage correct) and RT (in seconds) of the responses to the targets (averaged across familiar German words and unfamiliar English words). Error bars indicate 95 % confidence intervals. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

In order to test whether the children carried out the decision about the presence of a target just based on the onset of the word, we computed an additional analysis. We compared the false alarm rate to non-target words starting with a “p” (German word: PREIS; English word: PEN) to the false alarm rate of the remaining non-target words and to the accuracy of the targets across all trials using  $t$  tests. Accordingly, the children had a small, but significantly higher false alarm rate for the non-target P-words (2.3 %) than for the other non-target words (0.8 %;  $t(68) = 4.79$ ,  $p < 0.001$ ). However, the false alarm rate for the non-target P-words was far below the accuracy of the target detection (91.3 %;  $t(68) = -71.5$ ), indicating that the decisions were not just based on the onset of the words.

Moreover, we computed correlations between reading fluency and behavioral responses to visual targets (RT and accuracy), in order to test whether poorer reading skills were associated with slower and less accurate responses. We further computed analogous correlations for auditory targets to test whether such associations were specific for reading. Significant correlations were found for the visual modality (RT:  $r = -0.40$ ,  $p < 0.001$ ; accuracy:  $r = 0.40$ ,



**Fig. 3** A point-to-point comparison (TANOVA) to reveal time windows of audiovisual integration effects for familiar German words and unfamiliar English words. **Bold blue lines** indicate

significant time windows ( $p < 0.01$ ). Suppression effect (*upper row*): sumAV versus AVM. Congruency effect (*lower row*): AVN versus AVM (Color figure online)

$p < 0.001$ ), but not for the auditory modality (RT:  $r = -0.02$ ,  $p = \text{ns}$ ; accuracy:  $r = 0.18$ ,  $p = \text{ns}$ ).

#### Time Windows of Audiovisual Integration (Aim 1, First Step)

##### *Suppression Effects: sumAV versus AVM*

TANOVAs were applied across all 69 subjects separately for familiar German words and unfamiliar English words. As shown in Fig. 3 (upper row), the analyses revealed significant ( $p < 0.01$ ) suppression effects for familiar German words (300–324, 480–764 ms) and unfamiliar English words (324–384, 416–756 ms).

##### *Congruency Effects: AVN versus AVM*

The corresponding analyses to assess time windows of differential processing for matching and nonmatching audiovisual stimuli revealed significant ( $p < 0.01$ ) congruency effects only for familiar German words (160–204, 544–576, 1032–1108, 1164–1188 ms), but not for unfamiliar English words (see Fig. 3, lower row).

While these effects are illustrated in more detail in topographic maps below, the ERP waveforms are also illustrated for familiar German words at two exemplary channels (see Fig. 4).

#### Characterizing Time Segments (Aim 1, Second Step)

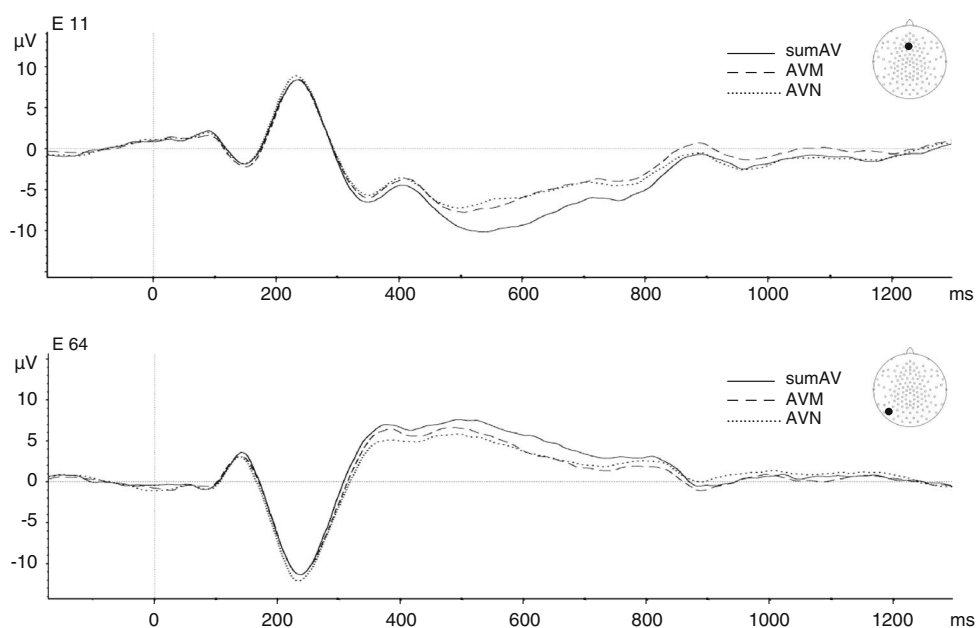
##### *Suppression Effects*

**Map Strength (GFP)** The ANOVAs on GFP for the four time segments (two time segments for familiar German words, two time segments for unfamiliar English words) identified by the TANOVA revealed significant main effects of suppression in all four time segments tested (see statistics in Table 1). As can be seen in Fig. 5, GFP was larger in the sumAV condition compared to the AVM condition in all four time segments.

**Map Topography (3D-Centroids)** The centroid analysis of the first time window of audiovisual suppression of familiar German words (300–324 ms) yielded a condition-by-polarity interaction ( $p < 0.001$ , see Table 1). This indicates that the centroids of the positivities and negativities were located differently between the summated auditory and visual conditions and the audiovisual matching condition (see Fig. 5). These topographic differences occurred mainly on the x-axis ( $p < 0.001$ ) and the z-axis ( $p < 0.05$ ), as revealed by planned comparisons.

Also for audiovisual suppression of unfamiliar English words a condition-by-polarity interaction ( $p < 0.01$ , see Table 1) was found in the first time window (324–384 ms). These topographic differences occurred mainly on the x-axis ( $p < 0.001$ ).



**Fig. 4** ERP waveforms for words at two exemplary channels**Table 1** Suppression effects

Suppression effect				
Time segment (ms)	GFP	Centroids (multivariate)		Significant centroid axes (univariate) <sup>a</sup>
Familiar German words				
300–324	F(1,68) = 39.236***	Condition	F(3,66) = 1.487	χ***, z*
		Condition × polarity	F(3,66) = 8.416***	
480–764	F(1,68) = 56.526***	Condition	F(3,66) = 0.197	
		Condition × polarity	F(3,66) = 0.337	
Unfamiliar English words				
324–384	F(1,68) = 17.153***	Condition	F(3,66) = 2.137	χ***
		Condition × polarity	F(3,66) = 4.804**	
416–756	F(1,68) = 39.876***	Condition	F(3,66) = 0.755	
		Condition × polarity	F(3,66) = 1.759	

<sup>a</sup> Only reported for significant multivariate effects

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

The centroid analysis of the second time windows (familiar German words: 480–764 ms; unfamiliar English words: 416–756 ms) of audiovisual suppression yielded no condition-by-polarity interaction, neither for familiar German words nor for unfamiliar English words (both  $p > 0.1$ ).

The main effect of condition did not reach significance in any of the time-windows, neither for familiar German words, nor for unfamiliar English words (all  $F < 2.2$ , for details see Table 1).

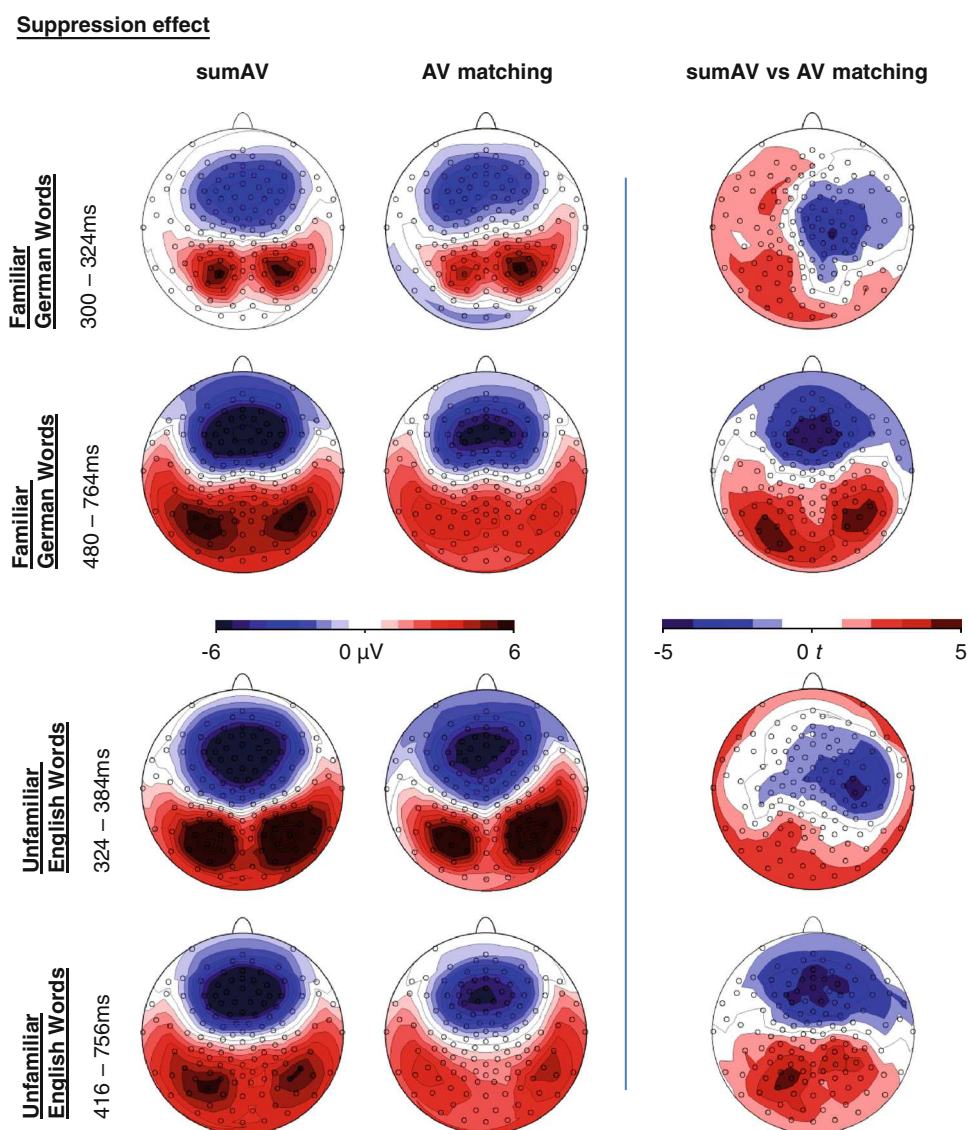
#### Congruency Effects

**Map Strength (GFP)** The ANOVAs on GFP for the four time segments identified by the TANOVA for familiar

German words revealed no significant main effects of congruency (see statistics in Table 2; compare Fig. 6). The effect in the 544–576 time window will not be interpreted, as it was only at a trend level.

**Map Topography (3D-Centroids)** The centroid analysis of all four time windows of audiovisual congruency of familiar German words yielded a condition-by-polarity interaction ( $p < 0.05$ ). The different centroid locations indicate different topographic distributions between audiovisual matching and audiovisual nonmatching conditions (see Fig. 6). Planned comparisons revealed that these topographic differences occurred mainly on the y-axis ( $p < 0.05$  in three of the four time segments). For

**Fig. 5** Topographic ERP maps and corresponding t-maps of the segments reflecting time windows of suppression effects (sumAV vs. AVM) for familiar German words and unfamiliar English words



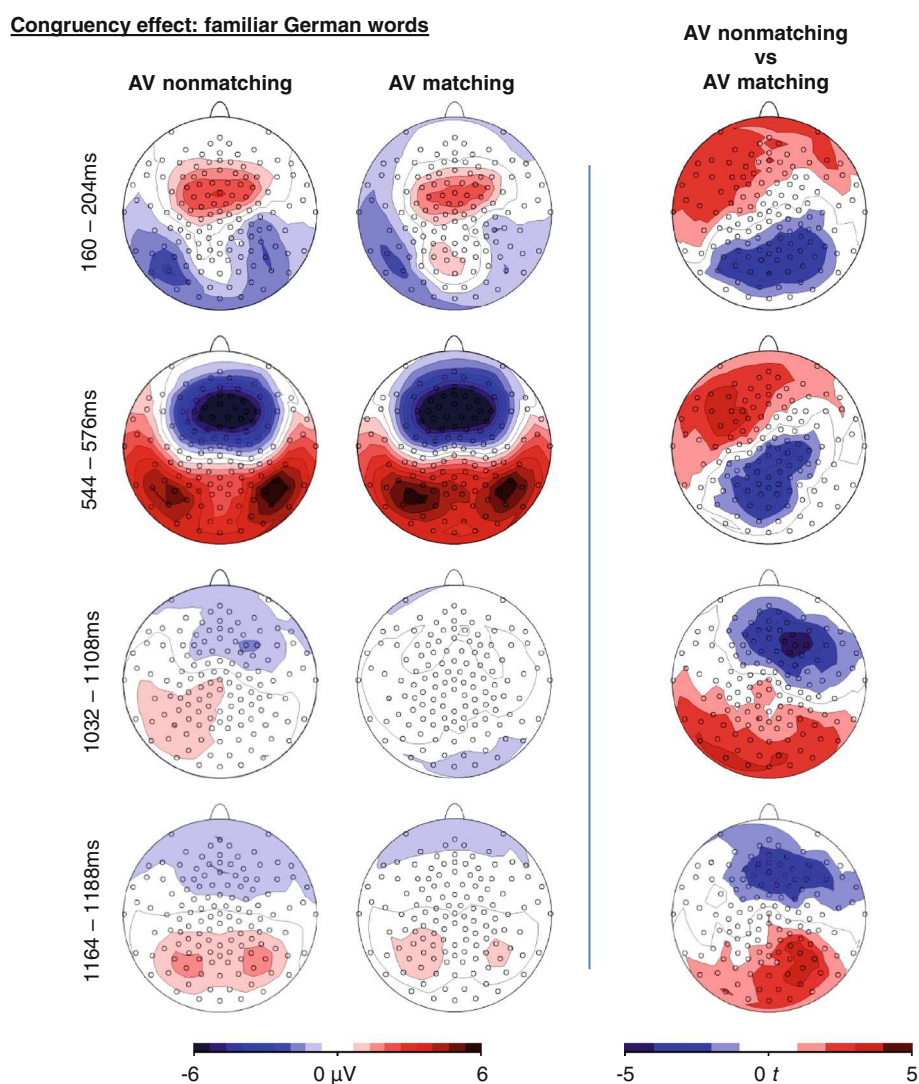
**Table 2** Congruency effects

Congruency effect familiar German words				
Time segment (ms)	GFP	Centroids (multivariate)		Significant centroid axes (univariate) <sup>a</sup>
160–204	$F(1,68) = 1.5$	Condition	$F(3,66) = 1.394$	
		Condition $\times$ polarity	$F(3,66) = 3.649^*$	$x^{\#}, y^*$
544–576	$F(1,68) = 3.647^{\#}$	Condition	$F(3,66) = 1.882$	
		Condition $\times$ polarity	$F(3,66) = 3.345^*$	$z^*$
1,032–1,108	$F(1,68) = 1.71$	Condition	$F(3,66) = 0.150$	
		Condition $\times$ polarity	$F(3,66) = 4.341^{**}$	$y^{**}$
1,164–1,188	$F(1,68) = .472$	Condition	$F(3, 66) = 0.632$	
		Condition $\times$ polarity	$F(3,66) = 3.804^*$	$y^{**}$

<sup>a</sup> Only reported for significant multivariate effects

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ ;  $\#$   $p < 0.1$  (trend)

**Fig. 6** Topographic ERP maps and corresponding t-maps of the segments reflecting time windows of congruency effects (AVN vs. AVM) for familiar German words



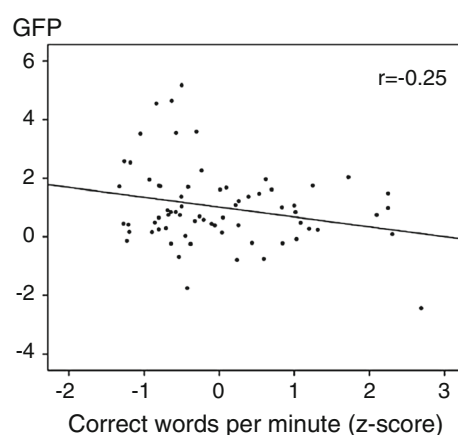
the second time window however, a topographic difference occurred on the z-axis ( $p < 0.05$ ).

The main effect of condition did not reach significance in any of the time-windows (all  $F < 1.9$ , for details see Table 2).

#### Association Between Reading Fluency and Audiovisual Integration (Aim 2)

In order to test whether audiovisual integration is associated with reading fluency we computed correlations between reading fluency and the differences between the two conditions (sumAV–AVM, AVN–AVM, respectively) that revealed significant GFP or centroid differences in the previous analysis.

Among the suppression and congruency effects reported, only the earlier suppression effect for familiar German words (300–324 ms, GFP) was significantly associated with



**Fig. 7** Scatterplot of reading fluency and GFP difference for the first time window reflecting suppression effects for familiar German words (300–324 ms). Reading fluency, defined as correct words per minute, was based on the average of 3 z-transformed subtests of the SLRT-I (Landerl et al. 2006) and SLRT-II (Moll and Landerl 2010)

reading skills ( $r = -0.246$ ,  $p < 0.05$ ). Visual inspection of the scatterplot in Fig. 7 suggested that the subject with the highest reading skills might have exerted undue influence. According to an outlier statistics, Cook's  $d$  of this subject was 0.30, which is below the threshold of 1, typically used for exclusion (Field, 2009). Yet, it was more than seven standard deviations above the mean of the Cook's  $d$  values of all subjects (mean: 0.016,  $\pm 0.04$  SD). Moreover, when excluding this subject the correlation was clearly not significant anymore ( $p > 0.19$ ). Thus, this correlation is not very robust and should be interpreted with caution.

None of the other reported suppression and congruency effects were significantly associated with reading fluency (all  $r < 0.18$ , all  $p > 0.148$ ).

Given the absence of robust associations between reading fluency and audiovisual integration measures, an additional question was whether reading fluency was associated with visual aspects of word processing. We thus computed correlations between reading fluency and GFP of the unimodal visual ERP in response to German words (P1, N1, P2, P3 components; as defined by grandmean GFP minima). None of these correlations were significant (all  $p$ 's  $> 0.2$ ). As associations between reading fluency and visual ERP components have been shown to be most pronounced for the N1 component at occipito-temporal electrodes in dyslexia studies (Maurer et al. 2007; Kast et al. 2010), we also computed correlations between left and right occipito-temporal electrode clusters and reading fluency. The correlation was significant for the left hemisphere ( $r = -0.33$ ,  $p < 0.01$ ; right hemisphere:  $r = 0.09$ ,  $p = \text{ns}$ ), indicating larger negativity of the N1 component in children who read more fluently.

## Discussion

Integrating visual and auditory language information is critical for reading. Yet, only little is known about how such multisensory processes relate to reading entire words and how they develop at the beginning of learning to read. The present study investigated audiovisual integration of familiar German words and unfamiliar English words in children after 1 year of reading training in school. Pursuing aim 1 of the study, we identified time windows of suppression (sumAV vs. AVM) and congruency (AVN vs. AVM) effects indicating audiovisual integration. In additional analyses as part of aim 1, we further characterized these effects on a neurophysiological level (GFP and topography). These effects, however, did not seem to be associated with reading skills (aim 2), presumably due to lack of automatization of reading mechanisms after 1 year of reading training. In the following, we discuss the results of aim 1 separately for suppression and congruency effects,

and combine results of the second aim for both effects in the end.

### Suppression Effects: sumAV versus AVM

TANOVAs revealed significant suppression effects for familiar German words (300–324, 480–764 ms) and unfamiliar English words (324–384, 416–756 ms). The earlier time window which was found for both familiar German words and unfamiliar English words concurs with the results from Raij et al. (2000), who also found audiovisual suppression effects around 350 ms. The similarity of the effects for familiar German words and unfamiliar English words is supported by the similar topographies, suggesting that similar processes are involved irrespective of the slightly different onset of these time segments. These processes thus seem to be unaffected by the lexical-semantic content of the stimuli. More likely sublexical processes are involved, e.g. grapheme-to-phoneme conversion or other phonological processes, possibly integrating print to speech irrespective of complexity of the stimulus, e.g., words, pseudowords, or single letters (Raij et al. 2000). This is also in agreement with ERP studies on visual word processing that demonstrated rhyming effects around 300–350 ms (Bentin et al. 1999; Spironelli and Angrilli 2007), suggesting phonological processing in this time range during reading.

In agreement with the results from Raij et al. (2000), the response to matching audiovisual stimuli in the earlier time window for both familiar German words and unfamiliar English words was smaller than the summed ERPs of the unimodal responses (sumAV  $>$  AVM), as revealed by differences in GFP. This suppressive effect of audiovisual presentation may reflect facilitation of the neural response during simultaneous audiovisual presentation of the same stimulus. When children learn to read, written and spoken words are often presented together and neural pathways that enable to memorize and retrieve audiovisual associations are formed. Consequently, the audiovisual suppression effect can be interpreted as an optimization of neural networks during learning (Raij et al. 2000).

The similar time range and the same direction of this effect (sumAV  $>$  AVM) in our study with children as in the Raij study with adults, may suggest that this early audiovisual suppression effect does not depend on the level of reading expertise. However, conclusions about the equivalence of the results in the two studies are difficult, as the two studies used different techniques (EEG vs. MEG) and different stimuli (single letters/phonemes vs. words). Moreover, it is possible that the different level of reading expertise in the two studies led to small but significant differences, which could only be detected in direct comparisons. Thus, developmental studies enabling contrasts between different age



groups with the same technique and experiment are needed to address this question more conclusively.

A facilitatory interpretation of the audiovisual integration effects is also supported by the behavioral results in the present study, which showed that children detected targets more accurately and faster when they were presented simultaneously in both modalities compared to when they were presented only visually or only auditorily (“multisensory facilitation”, e.g., Molholm et al. 2004).

Next to differences in GFP, the first time segment found for familiar German words and unfamiliar English words yielded differences in topography. Given that different topographies directly indicate differences in underlying sources (Michel et al. 2004), this result suggests that the simultaneous audiovisual presentation of familiar or unfamiliar words led to the activation of (partly) different brain regions compared to the summated activation in response to the corresponding unimodal auditory and visual stimuli.

The additive model (Raij et al. 2000) has been used in several studies investigating audiovisual interaction effects. Yet, this paradigm should be applied with caution as there are possible biases and artifacts that can be generated in the model (e.g., Besle et al. 2004, 2009; Teder-Salejarvi et al. 2002). A problem that can occur when applying the additive model is that neural activity common to all modalities (A, V, AVM) is added twice and subtracted only once: sumAV-AVM. It has been suggested that this problem can be avoided by for example restricting the analysis to an early time range, and also by avoiding motor responses in the signal to be analyzed (Besle et al. 2009). Our first time window occurred around 300–350 ms, which may be considered as early, taking into account the age of our participants and the complexity of the stimuli. In addition, only the target stimuli, which were not included in the ERPs, required any motor responses. Moreover, the different brain regions involved in this first audiovisual suppression effect, as indicated by the topographic differences, make it unlikely that neural activity common to all modalities was added twice and subtracted once. In such a case one would rather expect differences in GFP only (Cappe et al. 2010). Taken together, this suggests that the first suppression effect for familiar German words and unfamiliar English words truly reflects audiovisual integration processes, and is not merely an artifact resulting from the additive model.

In contrast to the first time window, the second time window (familiar German words: 480–764 ms; unfamiliar English words: 416–756 ms) of audiovisual suppression seemed to be reflected by differences in GFP only, indicating that same sources are active in the two conditions (sumAV and AVM), but that these sources are different in strength. Considering the possible artifacts that can be generated by the additive model (e.g., Besle et al. 2004, 2009; Teder-Salejarvi et al. 2002), the second time window of audiovisual suppression should be

interpreted with caution. Given that the differences were restricted to GFP and occurred in a late time range, this segment may not truly reflect audiovisual integration.

#### Congruency Effects: AVN versus AVM

TANOVAs revealed significant congruency effects only for familiar German words, but not for unfamiliar English words. The first time window during which matching and nonmatching words were processed differently occurred early (160–204 ms) in a time range which has been shown to reflect print processing in previous studies (e.g., Maurer et al. 2005, 2006). The presence of this effect for familiar German words but not for unfamiliar English words suggests that lexical-semantic information led to this early congruency effect. Although the occurrence of this effect appears to be early for semantic processing, early lexical-semantic modulations of visual word processing have been demonstrated previously in a series of studies (Hauk and Pulvermüller 2004; Sereno et al. 1998; Skrandies 1998; Wirth et al. 2007). The present study extends these studies by showing that such early lexical-semantic modulations also affect processing of audiovisual information.

The second time window showing a congruency effect occurred considerably later between 500 and 600 ms. The timing and the centroparietally negative topography of this effect are in agreement with N400 effects reported in previous studies (Kutas and Federmeier 2011 for a review; Schulz et al. 2008, 2009). The presence of this effect for familiar German words and its absence for unfamiliar English words is in agreement with the notion that it was driven by the semantic content of the words, as reported in N400 effects related to processing of meaning (Kutas and Federmeier 2011). This suggests that matching processes during audiovisual integration are similar to processes that are involved in semantic priming, and that these processes are restricted to meaningful stimuli.

The third and fourth time segment (1,032–1,108 ms; 1,164–1,188 ms) showed a congruency effect with posterior positivity. The occurrence of a late positivity after an N400 effect has often been reported in semantic priming studies (Van Petten and Luka 2012), and the topographic distribution is in agreement with previous findings (Schulz et al. 2008). A late positivity has not only been associated with syntactic but also semantic violations, and has been suggested to be triggered by integration difficulty and associated reprocessing costs (Van Petten and Luka 2012). The absence of this effect for unfamiliar English words in the present study suggests that meaningfulness is critical to elicit increased reprocessing of nonmatching audiovisual stimuli.

The ANOVAs on GFP for the 4 segments revealed no GFP differences, but topographic differences between matching and nonmatching words indicating that different



sources were activated in the brain depending on the congruency status. As opposed to previous studies (Calvert et al. 2000; Herdman et al. 2006; van Atteveldt et al. 2004), we did not find enhanced responses to congruent compared to incongruent audiovisual stimuli, nor did we find that activation was higher for incongruent than congruent stimuli (Hocking and Price 2009). In addition to differences in stimuli and neural measures (BOLD response, EEG frequencies) compared to those studies, a possible reason could be that our audiovisual experiment included an active task and it has previously been shown that the congruency effect disappears in active tasks, probably due to the fact that in decision tasks all stimuli pairs become equally relevant (van Atteveldt et al. 2007b). Moreover, GFP in the presence of topographic differences is less sensitive to measure differences in activation strength originating from specific brain regions.

#### Association Between Reading Fluency and Audiovisual Integration

Among all suppression and congruency effects reported above, only the first time window reflecting a suppression effect for familiar German words (300–324 ms) was found to be modulated by reading skills. The direction of the correlation indicates that the better a child read, the smaller the GFP difference between sumAV and AVM for words became. Although the opposite pattern (better reading fluency associated with a larger difference) might have been expected, the association found is in accordance with inverted u-shape patterns frequently found in developmental studies (Maurer et al. 2006): after the initial emergence this early audiovisual integration effect may become further reduced, when reading skills progress, possibly due to more efficient processing. Such an inverted u-shape pattern in the course of reading acquisition has also been shown for the development of neural tuning for print as indicated by the N1 ERP component (Maurer et al. 2006). Still, as we did not test the children at different time points during development, such developmental claims warrant further investigation.

There are two reasons, however, why the above correlation should be treated with caution. First, the correlation was found for the GFP measure in a segment, where also topographic effects were found, which indicate differences in underlying neural sources. A difference in activation strength (GFP) between different sources, is difficult to interpret given that some parts of the head are not covered by electrodes. Second, although the outlier statistics (Cook's  $d$  threshold) did not reveal any clear outliers, there was one case with extreme values in both variables which critically contributed to the correlation.

Given the theoretical importance of grapheme-phoneme conversion for learning to read (Blomert 2011; Wimmer

and Schurz 2010), a robust influence of reading skills on audiovisual integration mechanisms could have been expected. One reason for the absence of robust effects in any of the time windows reflecting audiovisual integration could be that in the topographic analyses we assumed a linear relation between centroid location and reading fluency. However, such a relation could also be non-linear, as larger topographic differences do not necessarily indicate a larger degree of audiovisual integration, although in some cases linear relations between topographic measures and behavioral measures can be found (Maurer et al. 2009).

Another potential factor affecting the sensitivity of the ERPs for modulation of audiovisual effects by reading ability was the large number of repetitions in the study. Stimulus repetition is known to reduce neural activity in various experimental paradigms (Grill-Spector et al. 2006). Such repetition suppression of the neural signal may not only reduce the chance to detect modulation of audiovisual interaction by reading ability, but also affect the detection of audiovisual interaction directly. Thus, the effects of audiovisual interaction reported in this study, may be seen as a conservative estimate of the presence of these effects in children at the beginning of learning to read. However, there is also evidence that semantic priming effects in the N400 are largely unaffected by stimulus repetition (Debruille and Renoult 2009). This suggests that the lexical-semantic congruency effects in the present study are less likely to be affected by the repetitions.

Moreover, there was also behavioral evidence that stimulus repetition did not override the effect of reading fluency in the current experiment. As could be expected, poorer readers detected the visual target word *PINK* more slowly and less accurately than more fluent readers.

Another reason for the absence of a robust association between reading fluency and audiovisual integration could be that our participants were first grade children with only 1 year of reading instruction when reading has not yet become automated. Possibly, modulation of audiovisual integration by reading skills at the beginning of learning to read might only be detected when using a more explicit reading task in the EEG. Indeed, the question to which degree the children explicitly read these words in the current study, cannot be fully answered. However, the low false alarm rate for the non-target words starting with the letter or the sound "P" compared to the high accuracy of the targets, as well as the absence of congruency effects for unfamiliar English words and their presence for familiar German words, suggest that lexical-semantic processing was involved.

Even though we did not find any robust associations between word reading fluency and measures of audiovisual integration, we found a significant correlation between reading fluency and the visual N1 component of the ERP in response to visual German words at left occipito-temporal electrodes (but

not in GFP). This may suggest that in the first year of learning to read, reading fluency is mainly reflected by visual aspects of word processing, but not in grapheme-phoneme conversion. However, we cannot exclude that such associations already exist in the first year of learning to read, and that we missed them in our analysis approach using global map descriptors (GFP, centroids) due to the lack of pre-existing knowledge that would allow to focus on critical regions.

## Conclusion

Given the novelty of the paradigm that uses entire words as well as the sample of children at an early stage of reading acquisition, this study provides new insights into the nature of audiovisual integration at the beginning of learning to read. First, we show that audiovisual integration takes place in children who had only 1 year of reading training and specify the time range when this integration occurs. Moreover, we show that audiovisual stimuli result in a suppressive neural response compared to the summated unimodal response, and that this suppressive effect is similar for familiar German words as for unfamiliar English words. Furthermore, audiovisually matching words elicit neural activation in different brain regions compared to audiovisually nonmatching words. This effect was only present for familiar German words and absent for unfamiliar English words, probably due to lexical-semantic processing. Moreover, audiovisual integration of words does not seem to be associated with reading fluency in the initial stages of reading acquisition, as no robust correlations were found in the present study. Thus, conclusions regarding the degree to which audiovisual suppression and congruency effects are associated with the level of reading skills, remains open to further investigation.

**Acknowledgments** This research was supported by the Swiss National Science Foundation (SNF-128610). We would also like to thank the two anonymous reviewers for their valuable input.

## References

- Achenbach T (1991) Manual for the child behavior checklist/4–18 and 1991 Profile. University of Vermont Department of Psychiatry, Burlington
- Baayen R, Piepenbrock R, Van Rijn H (1993) The CELEX lexical database [CD-ROM]. University of Pennsylvania Linguistic Data Consortium, Philadelphia
- Bentin S, Mouchetant-Rostaing Y, Giard MH, Echallier JF, Pernier J (1999) ERP manifestations of processing printed words at different psycholinguistic levels: time course and scalp distribution. *J Cogn Neurosci* 11(3):235–260
- Besle J, Fort A, Giard M-H (2004) Interest and validity of the additive model in electrophysiological studies of multisensory

- interactions. *Cogn Process* 5(3):189–192. doi:[10.1007/s10339-004-0026-y](https://doi.org/10.1007/s10339-004-0026-y)
- Besle J, Bertrand O, Giard M-H (2009) Electrophysiological (EEG, sEEG, MEG) evidence for multiple audiovisual interactions in the human auditory cortex. *Hear Res* 258(1–2):143–151. doi:[10.1016/j.heares.2009.06.016](https://doi.org/10.1016/j.heares.2009.06.016)
- Blau V, Reithler J, van Atteveldt N, Seitz J, Gerretsen P, Goebel R, Blomert L (2010) Deviant processing of letters and speech sounds as proximate cause of reading failure: a functional magnetic resonance imaging study of dyslexic children. *Brain* 133(Pt 3):868–879. doi:[10.1093/brain/awp308](https://doi.org/10.1093/brain/awp308)
- Blomert L (2011) The neural signature of orthographic–phonological binding in successful and failing reading development. *Neuroimage* 57(3):695–703. doi:[10.1016/j.neuroimage.2010.11.003](https://doi.org/10.1016/j.neuroimage.2010.11.003)
- Boersma P (2001) Praat: a system for doing phonetics by computer. *Glot Int* 5(9/10):341–345
- Calvert GA, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr Biol* 10(11):649–657
- Cappe C, Thut G, Romei V, Murray MM (2010) Auditory–visual multisensory interactions in humans: timing, topography, directionality, and sources. *J Neurosci* 30(38):12572–12580. doi:[10.1523/JNEUROSCI.1099-10.2010](https://doi.org/10.1523/JNEUROSCI.1099-10.2010)
- Debrulle JB, Renault L (2009) Effects of semantic matching and of semantic category on reaction time and N400 that resist numerous repetitions. *Neuropsychologia* 47:506–517
- Duyck W, Desmet T, Verbeke LPC, Brysbaert M (2004) WordGen: a tool for word selection and nonword generation in Dutch, English, German, and French. *Behav Res Methods, Instrum Comput* 36:488–499
- Ehri LC (2005) Learning to read words: theory, findings, and issues. *Sci Stud Read* 9(2):167–188. doi:[10.1207/s1532799xssr0902\\_4](https://doi.org/10.1207/s1532799xssr0902_4)
- Field A (2009) Discovering statistics using SPSS, 3rd edn. Sage, London
- Froyen D, Van Atteveldt N, Bonte M, Blomert L (2008) Cross-modal enhancement of the MMN to speech-sounds indicates early and automatic integration of letters and speech-sounds. *Neurosci Lett* 430(1):23–28. doi:[10.1016/j.neulet.2007.10.014](https://doi.org/10.1016/j.neulet.2007.10.014)
- Froyen DJ, Bonte ML, van Atteveldt N, Blomert L (2009) The long road to automation: neurocognitive development of letter–speech sound processing. *J Cogn Neurosci* 21(3):567–580
- Froyen D, Willems G, Blomert L (2011) Evidence for a specific cross-modal association deficit in dyslexia: an electrophysiological study of letter-speech sound processing. *Dev Sci* 14(4):635–648. doi:[10.1111/j.1467-7687.2010.01007.x](https://doi.org/10.1111/j.1467-7687.2010.01007.x)
- Giard MH, Peronnet F (1999) Auditory–visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J Cogn Neurosci* 11(5):473–490. doi:[10.1162/089892999563544](https://doi.org/10.1162/089892999563544)
- Grill-Spector K, Henson R, Martin A (2006) Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn Neurosci* 10(1):14–23. doi:[10.1016/j.tics.2005.11.006](https://doi.org/10.1016/j.tics.2005.11.006)
- Hauk O, Pulvermüller F (2004) Effects of word length and frequency on the human event-related potential. *Clin Neurophysiol* 115(5):1090–1103. doi:[10.1016/j.clinph.2003.12.020](https://doi.org/10.1016/j.clinph.2003.12.020)
- Herdman AT, Fujioka T, Chau W, Ross B, Pantev C, Picton TW (2006) Cortical oscillations related to processing congruent and incongruent grapheme–phoneme pairs. *Neurosci Lett* 399(1):61–66. doi:[10.1016/j.neulet.2006.01.069](https://doi.org/10.1016/j.neulet.2006.01.069)
- Hocking J, Price CJ (2009) Dissociating verbal and nonverbal audiovisual object processing. *Brain Lang* 108(2):89–96. doi:[10.1016/j.bandl.2008.10.005](https://doi.org/10.1016/j.bandl.2008.10.005)
- Holmes AP, Blair RC, Watson JD, Ford I (1996) Nonparametric analysis of statistic images from functional mapping experiments. *J Cereb Blood Flow Metab* 16(1):7–22

- Jung TP, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski TJ (2000) Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clin Neurophysiol* 111(10):1745–1758. doi:[S1388-2457\(00\)00386-2](https://doi.org/10.1016/j.clinph.2000.03.003)
- Kast M, Elmer S, Jancke L, Meyer M (2010) ERP differences of pre-lexical processing between dyslexic and non-dyslexic children. *Int J Psychophysiol* 77:59–69. doi:[10.1016/j.ijpsycho.2010.04.003](https://doi.org/10.1016/j.ijpsycho.2010.04.003)
- Klucharev V, Mottonen R, Sams M (2003) Electrophysiological indicators of phonetic and non-phonetic multisensory interactions during audiovisual speech perception. *Cogn Brain Res* 18(1):65–75. doi:[10.1016/j.cogbrainres.2003.09.004](https://doi.org/10.1016/j.cogbrainres.2003.09.004)
- Kutas M, Federmeier KD (2011) Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Ann Rev Psychol* 62:621–647. doi:[10.1146/annurev.psych.093008.131123](https://doi.org/10.1146/annurev.psych.093008.131123)
- Landerl K, Wimmer H, Moser E (2006) SLRT Salzburger Lese- und Rechtschreibtest (Salzburg reading and orthography test). Huber, Bern
- Lehmann D, Skrandies W (1980) Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalogr Clin Neurophysiol* 48(6):609–621
- Maurer U, Bucher K, Brem S, Brandeis D (2003) Development of the automatic mismatch response: from frontal positivity in kindergarten children to the mismatch negativity. *Clin Neurophysiol* 114(5):808–817. doi:[10.1016/S1388-2457\(03\)00032-4](https://doi.org/10.1016/S1388-2457(03)00032-4)
- Maurer U, Brem S, Bucher K, Brandeis D (2005) Emerging neurophysiological specialization for letter strings. *J Cogn Neurosci* 17(10):1532–1552
- Maurer U, Brem S, Kranz F, Bucher K, Benz R, Halder P, Steinhausen H-C, Brandeis D (2006) Coarse neural tuning for print peaks when children learn to read. *Neuroimage* 33(2):749–758. doi:[10.1016/j.neuroimage.2006.06.025](https://doi.org/10.1016/j.neuroimage.2006.06.025)
- Maurer U, Brem S, Bucher K, Kranz F, Benz R, Steinhausen H-C, Brandeis D (2007) Impaired tuning of a fast occipito-temporal response for print in dyslexic children learning to read. *Brain* 130:3200–3210. doi:[10.1093/brain/awm193](https://doi.org/10.1093/brain/awm193)
- Maurer U, Bucher K, Brem S, Benz R, Kranz F, Schulz E, van der Mark S, Steinhausen H-C, Brandeis D (2009) Neurophysiology in preschool improves behavioral prediction of reading ability throughout primary school. *Biological Psychiatry* 66(4):341–348
- Maurer U, Rossion B, McCandliss BD (2008) Category specificity in early perception: face and word n170 responses differ in both lateralization and habituation properties. *Front Hum Neurosci*. doi:[10.3389/neuro.09.018.2008](https://doi.org/10.3389/neuro.09.018.2008)
- Maurer U, Blau VC, Yoncheva YN, McCandliss BD (2010) Development of visual expertise for reading: rapid emergence of visual familiarity for an artificial script. *Dev Neuropsychol* 35(4):404–422. doi:[10.1080/87565641.2010.480916](https://doi.org/10.1080/87565641.2010.480916)
- Mayringer H, Wimmer H (2003) Salzburger Lese-Screening für die Klassenstufen 1–4. Manual. Hans Huber, Bern
- Michel CM, Murray MM, Lantz G, Gonzalez S, Spinelli L, Grave de Peralta R (2004) EEG source imaging. *Clin Neurophysiol* 115(10):2195–2222. doi:[10.1016/j.clinph.2004.06.001](https://doi.org/10.1016/j.clinph.2004.06.001)
- Mittag M, Thesleff P, Laasonen M, Kujala T (2013) The neurophysiological basis of the integration of written and heard syllables in dyslexic adults. *Clin Neurophysiol* 124:315–326. doi:[10.1016/j.clinph.2012.08.003](https://doi.org/10.1016/j.clinph.2012.08.003)
- Molholm S, Ritter W, Javitt DC, Foxe JJ (2004) Multisensory visual-auditory object recognition in humans: a high-density electrical mapping study. *Cereb Cortex* 14(4):452–465. doi:[10.1093/cercor/bhh007](https://doi.org/10.1093/cercor/bhh007)
- Moll K, Landerl K (2010) SLRT-II: Lese- und Rechtschreibtest. Weiterentwicklung des Salzburger Lese- und Rechtschreibtests (SLRT). Verlag Hans Huber, Bern
- Murray MM, Spierer L (2009) Auditory spatio-temporal brain dynamics and their consequences for multisensory interactions in humans. *Hear Res* 258(1):121–133. doi:[10.1016/j.heares.2009.04.022](https://doi.org/10.1016/j.heares.2009.04.022)
- Murray MM, Michel CM, Grave de Peralta R, Ortigue S, Brunet D, Gonzalez Andino S, Schneider A (2004) Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. *Neuroimage* 21(1):125–135. doi:[10.1016/j.neuroimage.2003.09.035](https://doi.org/10.1016/j.neuroimage.2003.09.035)
- Petermann F, Petermann U (2010) HAWIK-IV. Hamburg-Wechsler-Intelligenztest für Kinder—IV (Wechsler intelligence scale for children (WISC-IV)—German version). Huber, Bern
- Raij T, Uutela K, Hari R (2000) Audiovisual integration of letters in the human brain. *Neuron* 28(2):617–625
- Schulz E, Maurer U, van der Mark S, Bucher K, Brem S, Martin E, Brandeis D (2008) Impaired semantic processing during sentence reading in children with dyslexia: combined fMRI and ERP evidence. *Neuroimage* 41(1):153–168. doi:[10.1016/j.neuroimage.2008.02.012](https://doi.org/10.1016/j.neuroimage.2008.02.012)
- Schulz E, Maurer U, van der Mark S, Bucher K, Brem S, Martin E, Brandeis D (2009) Reading for meaning in dyslexic and young children: Distinct neural pathways but common endpoints. *Neuropsychologia* 47:2544–2557. doi:[10.1016/j.neuropsychologia.2009.04.028](https://doi.org/10.1016/j.neuropsychologia.2009.04.028)
- Sereno SC, Rayner K, Posner MI (1998) Establishing a time-line of word recognition: evidence from eye movements and event-related potentials. *NeuroReport* 9(10):2195–2200
- Skrandies W (1998) Evoked potential correlates of semantic meaning—a brain mapping study. *Cogn Brain Res* 6(3):173–183
- Spironelli C, Angrilli A (2007) Influence of phonological, semantic and orthographic tasks on the early linguistic components N150 and N350. *Int J Psychophysiol* 64(2):190–198. doi:[10.1016/j.ijpsycho.2007.02.002](https://doi.org/10.1016/j.ijpsycho.2007.02.002)
- Stein M, Dierks T, Brandeis D, Wirth M, Strik W, Koenig T (2006) Plasticity in the adult language system: a longitudinal electrophysiological study on second language learning. *Neuroimage* 33(2):774–783. doi:[10.1016/j.neuroimage.2006.07.008](https://doi.org/10.1016/j.neuroimage.2006.07.008)
- Stock C, Marx P, Schneider W (2003) BAKO 1–4. Basiskompetenzen für Lese-Rechtschreibleistungen, Beltz
- Strik WK, Fallgatter AJ, Brandeis D, Pascual-Marqui RD (1998) Three-dimensional tomography of event-related potentials during response inhibition: evidence for phasic frontal lobe activation. *Electroencephalogr Clin Neurophysiol* 108(4):406–413
- Teder-Salejarij WA, McDonald JJ, Di Russo F, Hillyard SA (2002) An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. *Cogn Brain Res* 14(1):106–114
- van Atteveldt N, Formisano E, Goebel R, Blomert L (2004) Integration of letters and speech sounds in the human brain. *Neuron* 43(2):271–282
- van Atteveldt NM, Formisano E, Blomert L, Goebel R (2007a) The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cereb Cortex* 17(4):962–974. doi:[10.1093/cercor/bhl007](https://doi.org/10.1093/cercor/bhl007)
- van Atteveldt NM, Formisano E, Goebel R, Blomert L (2007b) Top-down task effects override automatic multisensory responses to letter-sound pairs in auditory association cortex. *Neuroimage* 36(4):1345–1360. doi:[10.1016/j.neuroimage.2007.03.065](https://doi.org/10.1016/j.neuroimage.2007.03.065)
- Van Petten C, Luka BJ (2012) Prediction during language comprehension: benefits, costs, and ERP components. *Int J Psychophysiol* 83(2):176–190. doi:[10.1016/j.ijpsycho.2011.09.015](https://doi.org/10.1016/j.ijpsycho.2011.09.015)
- Wimmer H, Schurz M (2010) Dyslexia in regular orthographies: manifestation and causation. *Dyslexia* 16(4):283–299. doi:[10.1002/dys.411](https://doi.org/10.1002/dys.411)
- Wirth M, Horn H, Koenig T, Stein M, Federspiel A, Meier B, Michel C, Strik W (2007) Sex differences in semantic processing: event-related brain potentials distinguish between lower and higher order semantic analysis during word reading. *Cereb Cortex* 17(9):1987–1997. doi:[10.1093/cercor/bhl121](https://doi.org/10.1093/cercor/bhl121)